理工学専攻修士論文要旨

提 出 年 度： 2005 年度
提　 出　 日： 2006 年 2 月 27 日
専 修 コ ー ス：知能情報コース
学 生 番 号： 35604177
学 生 氏 名： 村上　裕人
研究指導教員：原田　実 教授

（論文題目）
自由記述アンケート文の自動分類システム AQUA の分類精度向上

（内容の要旨）
自由記述アンケートは選択型アンケートに比べ回答者の自由な意見を集約できるという効果があるため社会的にも注目されている．自由記述アンケートの分類などに応用が考えられる既存のテキスト自動分類では文の頻度ベクトル間のコサイン距離による手法や表層格の類似性に基づく手法が提案されているが，これらは意味解析を行っていないので，表層のゆれを吸収できず意味的な類似性による分類ができない。そこで，2003 年度原田研究室では，意味解析による自由記述アンケートの自動分類システムAQUA2003 の開発を行った．しかし，従来のAQUAでは，対象を意見の中心となる述語とそれに係る名詞節に限定したため，修飾やその他の語の違いを表現できないという問題点や最適なクラスタ数を得るためのコストが大きいという問題点があった．

そこで本研究では，AQUAの分類精度とユーザビリティの向上を目的とし，１）意味グラフの照合結果を用いた文全体を考慮した文類似度の提案，２）最適なクラスタ数の自動決定，３）クラスタ要約生成機能の追加，を行った．

１）の文類似度は，基本的に回答文の意味解析結果である２つの意味グラフのノードの語意類似性とアークの格類似性に基づく照合により得られた照合ノードペア類似度和であるノード類似度と照合アークペア類似度和であるアーク類似度の和であるが，意見分類では，「何をどうしたい」といった語間の深層格を含んだ類似性が重要であると考え，ノード類似度よりも照合アークペアとそのアークの両端ノードを重視する文類似度を提案した。また，話し手の判断・態度を表すムードをグルーピングし，願望や要求を表すグループのムードが付与された照合ノードペアの照合ノードペア類似度とそのノードを両端に持つ照合アークペア類似度に対して特に高得点を与える文類似度とした．２）では分類指数と呼ぶ統計量を３つ用意し，それらの数値が示すクラスタ数の平均値を最適なクラスタ数とした．この分類指数は，i）Gap 統計量，ii) Split 統計量，iii）Global 統計量の３つである．i）Gap 統計量は既存研究で提案された手法で，分析データと一様に分布するデータの間でクラスタ内のばらつきの差が最も大きくなるクラスタ数を最適とする．ii）Split 統計量は，あるデータ対（文のペア）が同クラスタまたは他クラスタに分類された際のクラスタリングの良さを文類似度で表現し，全データ対での合計値が最大値となるとき最適なクラスタ数とする．iii）Global 統計量はあるクラスタリング状態においてあるデータから見た局所的な複雑さの最大値と全体的な複雑さの和で表すものでこの２つはトレードオフの関係にあり，最小値となるとき最適なクラスタ数とする．３）のクラスタ要約の生成ではラベル型要約と整文型要約の 2 つを用意する。まず，意味グラフの照合によって得られた照合ノードペア類似度と照合アークペア類似度によりクラスタ内の文の語句に得点付けをし，重要語句選定をする．そして，ラベル型要約では，重要語句のみを要約として出力し，整文型要約では重要語句から main 格までのアークパス上の語全てを要約として出力する．

本システム AQUA2005 と従来システム AQUA2003 について，人手による分類結果との類似度合いの F 値を算出すると本システムでは約 0.73, 従来システムでは約 0.61 となり、システムの改善が見られた。また、クラスタ数の決定コストが減少し、クラスタ内要約により同クラスタ内のアンケート回答の分析支援が可能となった。